(2)

Naval Oceanographic and
Atmospheric Research Laboratory

Technical Note 27
May 1990

Compaction of Global Data Fields

A. H. Van Tuyl
National Center for Atmospheric Research
Boulder, CO 80307-3000

Prepared for
Prediction Systems Division
Atmospheric Directorate
Monterey, CA 93943-5006

ABSTRACT

Two methods of compacting global data fields are studied, both individually and in combination, and errors associated with the methods are systematically examined. The first scheme consists of expanding the data into an empirical orthogonal function (EOF) series in the vertical, then truncating the series at a selected number of terms. Because the EOFs are ordered by decreasing variance explained, this reduces the number of degrees of freedom while retaining most of the important vertical structure information. The second technique used is bit reduction, in which appropriately scaled data (here, spectral coefficients) are converted to integer form, with the scaling factor chosen so that the maximum data value is the largest integer expressible by some desired number of bits. Examination of compaction errors for various EOF truncations and bit scalings indicates that one important result of bit reduction is to set to zero all coefficients with magnitude below a certain threshold, causing EOF truncation up to a given point to have no impact on errors. Based on a somewhat arbitrarily selected maximum allowable RMS temperature error of 1^{o}C, a compaction factor of approximately two is obtainable from EOF truncation alone, and an additional factor of three from bit reduction (32 bits to 10), assuming half precision words. Future work is needed, however, to determine the generality of these results.

ACKNOWLEDGMENTS

The author wishes to acknowledge the support of the sponsor, J. Cauffman, code 22, Office of Naval Technology, Program Element 62435N, for making this effort possible.

TABLE OF CONTENTS

COMPACTION OF GLOBAL DATA FIELDS

1. INTRODUCTION

The objective of this study was to examine the impact of two data compaction methods on the information content of the data. These compaction schemes are vertical empirical orthogonal function (EOF) truncation, and bit truncation of spectral coefficients. The two methods were tested individually and in combination. Loss of information content was evaluated by comparing grid point fields computed from the spectral coefficients before and after compaction. The original coefficients were obtained from a T47 spherical harmonic truncation of the data, this truncation being unchanged by the compaction techniques; therefore, error due to horizontal resolution was not considered. Error statistics were evaluated globally, and, in some instances, over the combined North Atlantic and North Pacific regions. Additionally, in the latter cases the effect of redefining the EOF's was examined; that is, rather than being computed from global data they were computed only from data within the regions of interest.

2. BACKGROUND

The spectral truncation employed by the NOGAPS model (from which the analyses in this report were obtained) results in a four to one compaction of gridded data. Specifically, the use of a triangular, rather than rhomboidal, truncation accounts for a compaction factor of approximately two, with the remaining factor of two resulting from removal of the smallest scales which is done to avoid aliasing. (Approximately two-thirds of the wavenumbers in a given direction are retained.) In the present study, we examine methods of compressing the data still further, specifically excluding any additional truncation in the horizontal. Therefore, the cases discussed in this paper use the four to one compaction, and, most significantly, the errors shown do not include the errors of this truncation. The compaction methods which we do consider, and for which we evaluate errors, are (1) vertical empirical orthogonal function (EOF) truncation, and (2) bit truncation of spectral coefficients. These are examined individually below.

2.1 Vertical EOF Truncation

Empirical orthogonal functions (EOF's) are defined as the eigenvectors of a variance/ covariance matrix. The eigenvalue corresponding to each eigenvector has amplitude proportional to the amount of variance explained by that eigenvector. In this investigation EOF's are used as vertical structure functions, the covariance matrix being computed from correlations of a variable between different atmospheric levels. The leading eigenvectors (i.e., those with the largest eigenvalues) then represent the atmosphere's dominant vertical structures in that variable. Because, in

general, there is a wide range in magnitude between the largest and smallest eigenvalues. one may truncate the EOF expansion and still retain most of the significant information. This in turn becomes a source of data compaction, as the number of EOF's in the full expansion is the same as the number of levels.

2.2 Bit Truncation of Coefficients

The series of spectral coefficients consists of a set of complex numbers. Further compaction may be obtained by scaling the coefficients and converting them to integers, which allows the largest coefficients to be expressed by a specified maximum number of bits. The compressed data then consists of a set of integers together with a scaling factor, which is an integer J divided by the largest coefficient magnitude. J is selected to be the maximum number expressible in terms of some chosen number of bits. Because both positive and negative coefficients are in general present, note that it is also necessary to reserve one bit for the sign. The value of J, therefore, assuming a desired maximum of M bits, is given by

J=2**(M-1)-1.

For purposes of this study the coefficients are left in complex form (that is, not converted to amplitude and phase); the maximum value used for scaling is then either the real part or the imaginary part of some particular coefficient. All other real and imaginary parts are scaled by this single value. Scaled coefficients are rounded to the nearest integer, rather than truncated after the decimal point, in order to reduce loss of accuracy.

Upon rescaling the coefficients to their original magnitudes, the maximum coefficient value described above is recovered exactly (neglecting any floating-point errors). Other values, however, lose varying amounts of precision due to the integer conversion. This loss is proportionally greater for the smaller coefficients; indeed, those coefficients with scaled values of less than 0.5 become identically zero. Due to the fact that there is usually a difference of many orders of magnitude between the largest and smallest coefficients, the information content of a great number of coefficients is completely lost in the compaction process. In particular, the "red-noise" nature of many atmospheric and oceanic spectra, with energy concentrated at the largest scales, means that small-scale information is preferentially removed. Although one could argue that the eliminated coefficients are unimportant by definition, since they would not be eliminated in the first place if they contributed substantially to the total variance, it is not clear that this criterion is the only one by which to judge their value. For example, the coefficients corresponding to smaller scales may be important in resolving features such as fronts which, being small-scale and isolated, have little influence on the total (domain integrated) spectral amplitude. Also, should compacted fields be used to compute quantities involving horizontal spatial derivatives, particularly higher-order ones,

the significance of the small-scale information becomes relatively greater. One method of overcoming the above difficulty would be to scale deviations from some background spectrum resembling that of the atmosphere; for efficiency it would of course be necessary that this spectrum be represented by relatively few parameters. A piecewise linear function of wavenumber, for example, could be used. For simplicity this technique was not tried in the present investigation.

2.3 Source and Nature of Data

The source of all data used in this study is a single three-dimensional, global, grid-point analysis of temperature, valid 1200Z 19 April 1989, from the NOGAPS global atmospheric model. Horizontal coordinates are latitude and longitude, with a 2.5 degree resolution in both directions; pressure is the vertical coordinate. Eleven vertical levels are employed, ranging from 1000 to 50 millibars. The field was computed originally from spectral coefficients on sigma-surfaces, then interpolated to the surfaces of constant pressure; consequently, some vertical interpolation error is present. For this reason, the input grid is used to calculate spherical harmonic coefficients at each level, with T47 truncation (the same as in the NOGAPS 3.1 model); subsequently the inverse transform is applied. The resulting grid is the control case against which errors are evaluated. This procedure is followed because we desire to examine only the increase in error beyond that caused by the vertical interpolation scheme.

2.4 Analysis Regions

Errors are evaluated over an analysis region consisting of all levels in the vertical, and either a global or limited-area domain in the horizontal. The limited-area domain consists of the North Atlantic and North Pacific regions combined, and is thus actually comprised of two unconnected subdomains. Latitude and longitude coordinates for the North Pacific subdomain are 0 to 50N, 155E to 125W, and for the North Atlantic 5N to 60N, 55W to 17.5W. These values are chosen to enclose as much of the northern oceans as possible while excluding virtually all land areas, so that errors due to topographic influence will be negligible. As previously stated the ocean-only regions are also used for EOF computation in certain cases, in addition to error analysis.

2.5 Software

Most of the software employed in this study was written by various persons at NOARL-West or FNOC, including some routines written by this investigator. The only exceptions are the fast Fourier transform (FFT) software used in the spherical harmonic computations, and the matrix eigenvector routine for computing EOF's, which were obtained originally from NCAR. All of the software from the different sources was combined into one program by the investigator.

3. PROCESSING

As previously mentioned the two compaction techniques were studied individually and in combination. This section describes in more detail the specific procedures used in applying them to the data.

3.1 EOF Truncation

The three-dimensional input grid discussed in Section 2.4 was read from disk, stored in an array within the main program, and the horizontal mean at each level then removed. From the resulting array an 11 by 11 variance/covariance matrix was computed, with each element representing the correlation between two vertical levels (including correlations of a level with itself). The correlations were summed over all horizontal grid points in the analysis domain; thus the matrix contained only vertical dependence. (This vertical dependence, of course, was a function of the data characteristics within the particular horizontal analysis region.) From the covariance matrix, eigenvalues and eigenvectors were calculated using a standard routine, and the eigenvectors normalized. The eigenvector matrix was then inverted (equivalent to transposition in this case), and the resulting array used to project the vertically discretized data onto the EOF's. This projection was performed for every horizontal grid point, so that for each such point there then existed a vector of 11 EOF coefficients. The number of degrees of freedom was unchanged by this procedure, with the 11 vertically discrete values (for each horizontal point) merely replaced by the 11 coefficients.

After the EOF coefficients were computed, a spherical harmonic transform was applied to each of the 11 horizontal coefficient arrays; that is, the coefficients were treated as ordinary gridded data. The result was a set of coefficients with three mode indices (one for each spatial dimension). For the control case, this coefficient set was transformed back to grid space in the horizontal and vertical directions and the horizontal means restored. This is mathematically equivalent to performing the spherical harmonic transform and inverse transform at each level in physical space, since no EOF truncation was done in this instance. (A comparative experiment indicated that the computed difference between the two procedures was indeed negligible.) For the experiments with EOF truncation, all coefficients corresponding to the first N EOF's were set to zero, where N was an integer between 1 and 10 and the EOF's were ordered by increasing eigenvalue (the eigenvector routine performs this ordering automatically). Then the horizontal and vertical inverse transforms (and restoration of means) were applied as before. Finally, the resulting three-dimensional grid was compared with the control case and error statistics for the difference field (defined as truncated minus control) were calculated.

3.2 Bit Truncation

Bit truncation was performed on the three-dimensional (i.e., triply indexed) spectral coefficients for experiments with and without EOF truncation. In the case of no EOF truncation the method of Section 2.2 was followed exactly. When fewer than 11 EOF's were retained, the technique was to first zero out the unwanted coefficients, then perform the scaling on the coefficients which were nonzero. This was done to avoid the possibility of scaling by a coefficient which would later be set to zero, thus resulting in a more severe truncation of the remaining coefficients than necessary. (Since the coefficients eliminated were those of the less important EOF's, however, the likelihood of such a difficulty occurring was probably small.) After bit scaling and/or EOF truncation, the coefficients were used to compute a three-dimensional grid field and this field compared with the control case, as before.

4. RESULTS

Experiments were performed for every value of N (the number of EOF's removed) between 0 and 10, and for maximum bit values of 9, 10, and 11. (These bit values correspond to maximum scaled coefficients of 255, 511, and 1023, respectively.) Also, EOF truncation was examined in the absence of bit reduction. In this section we present certain of the more significant results in detail. A general summary of the remaining cases is also given, but we do not attempt to discuss each one individually.

4.1 General Results

Plots of maximum absolute error, average absolute error, and RMS error (all global) as a function of N were constructed for each of the four bit scaling cases (i.e., the non-scaled case plus the three scaled ones). Examination of the error curves (presented for the RMS values in Fig. 1) shows a number of significant features. First, as one would expect, the RMS and average absolute errors increase as fewer bits and fewer EOF's are retained. Interestingly, this is generally but not always true for the maximum absolute error, although any exceptions are small in magnitude. The differing behavior of this quantity is most likely due to the point value (rather than domain-averaged value) that it represents, resulting in greater variability between cases. A second characteristic of interest is that, as bit truncation is increased, the increase in error with removal of EOF's diminishes; that is, curves corresponding to more severe bit truncation are more "flat". This result is caused by the bit truncation which rounds the smaller coefficients to zero. Specifically, for a given bit truncation, there is a certain minimum number of zero coefficients (and a certain minimum number of EOF's for which all coefficients are zero); removal of coefficients up to that minimum then has no effect. For example, in the case of 9 maximum retained bits, Fig. 1 indicates

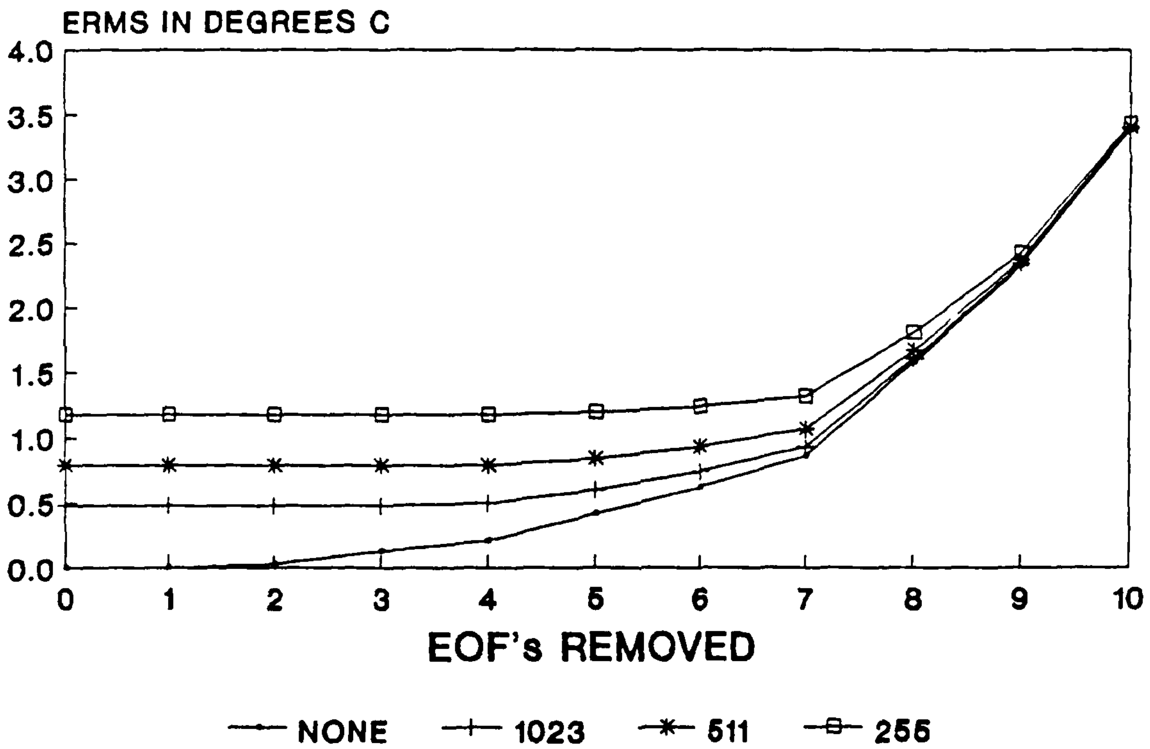

Fig. 1. Globally-evaluated root-mean-square temperature errors (degrees Centigrade) as a function of N (number of vertical EOF's omitted) for: no bit reduction, maximum of 11 bits (1023), maximum of 10 bits (511), maximum of 9 bits (255).

that the RMS error does not show an appreciable increase until at least 6 or 7 EOF's, more than half, have been removed.

As the number of EOF's retained is decreased, the difference among the bit scaling cases decreases also, and the four curves gradually converge. When only 1 EOF is retained the errors for all cases are virtually identical. Thus, for mild EOF truncation (most EOF's retained) the bit reduction is the dominant source of error, while for severe EOF truncation, the effect of removing EOF's is most important. This again confirms that the main result of the bit truncation is to zero out a number of the smaller coefficients, and that this number represents a threshold value beyond which EOF truncation becomes significant. In an operational situation, of course, it is not likely that only a few vertical EOF's would be retained. Still, it is of interest to note that the error introduced by retaining only the one most dominant EOF is much less than the error caused by eliminating only that EOF. For the present analysis at least, it therefore appears that the leading EOF's do in fact include most of the important vertical structure information.

Magnitudes of maximum absolute errors for the case of no EOF truncation are (identically) zero, 2.7, 4.1, and 6.7 degrees Centigrade, in order of increasing bit truncation. When only 2 EOF's are retained the corresponding error is about 17 degrees Centigrade, independent of bit reduction. (This error slightly decreases when an additional EOF is removed, for reasons discussed previously.) The RMS errors (Fig. 1) for no EOF truncation and the four respective bit scalings are 0, 0.5, 0.8, and 1.2 degrees C. Again the errors become identical for the most severe EOF truncation, and are equal to 3.4 degrees C when just the leading EOF is retained. Average absolute errors behave similarly to the RMS errors, but their values are approximately 25 percent smaller. Note that for the strongest bit truncation, and certainly for the maximum EOF truncation, the errors appear unacceptably high. For this reason we do not consider such cases further; their inclusion here is intended only to give a general idea of the characteristics of our compaction methods.

4.2 Effects of Varying Analysis Region

In this subsection we examine certain of the above results in more detail, and also investigate the effect that the choice of analysis domain has on error magnitude. As previously mentioned the influence of the analysis region is studied both from the standpoint of error evaluation (i.e., different areas may have different error characteristics), and in terms of the dependence of vertical EOF structure on the data within the region. Specific results are presented in Fig. 2 a-d, which respectively show error curves for each of the four bit scaling cases. Only examples with N=4,5,6 are considered. Figure 2a depicts, for the no-scaling case, RMS temperature errors evaluated globally (G-G), over the northern oceans (G-L), and over the northern oceans with EOF's computed from data in these regions (L-L). The most significant feature to note is that

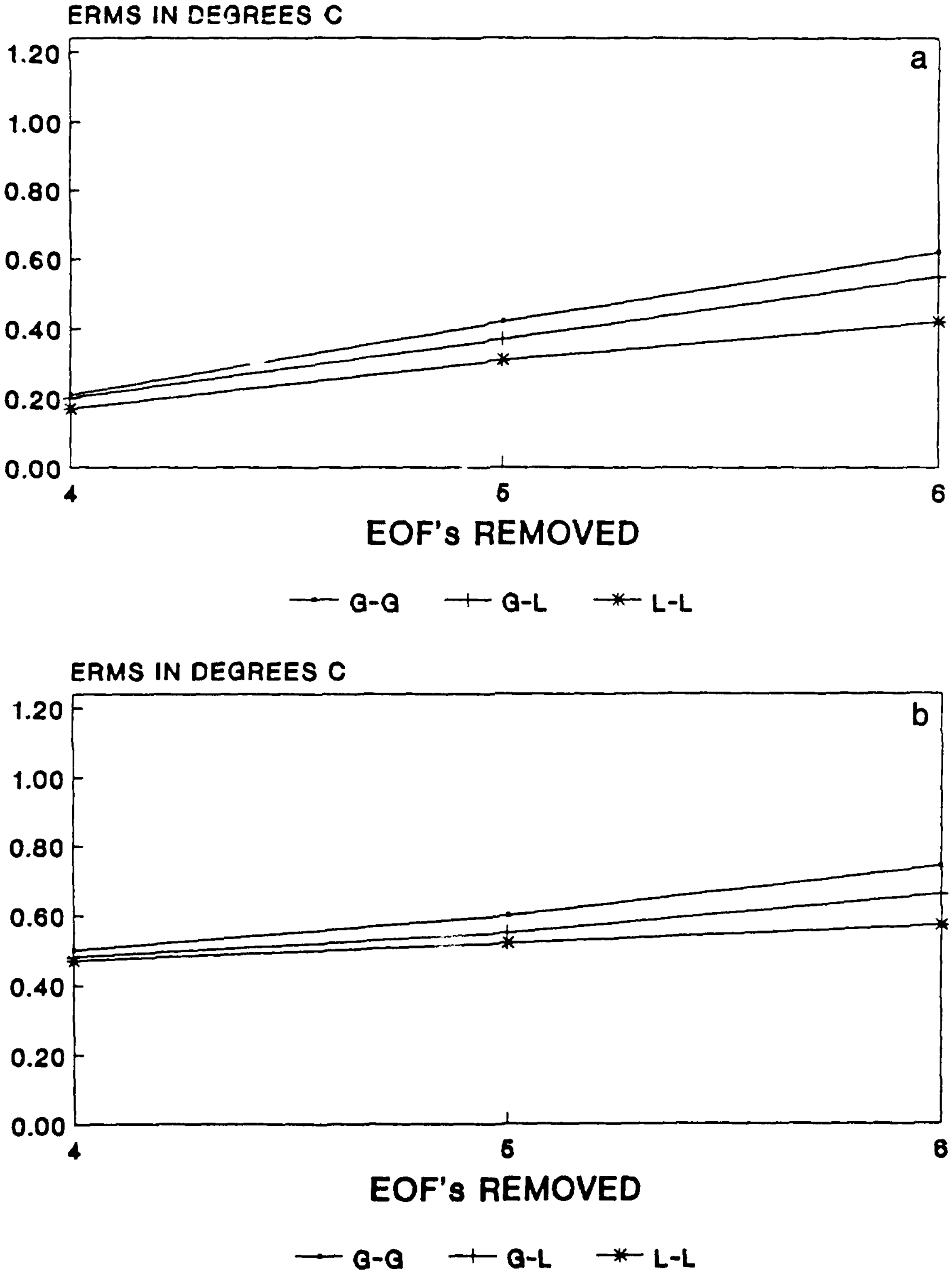

Fig. 2. Root-mean-square temperature errors (degrees Centigrade) as a function of N, evaluated globally with globally-defined EOF's (G-G), over the northern oceans with globally-defined EOF's (G-L), over the northern oceans with EOF's defined using data from these regions (L-L), for: (a) no bit truncation, (b) truncation to a maximum of 11 bits, (c) truncation to a maximum or 10 bits, (d) truncation to a maximum of 9 bits. Note that the temperature scale is different from that of Fig. 1.

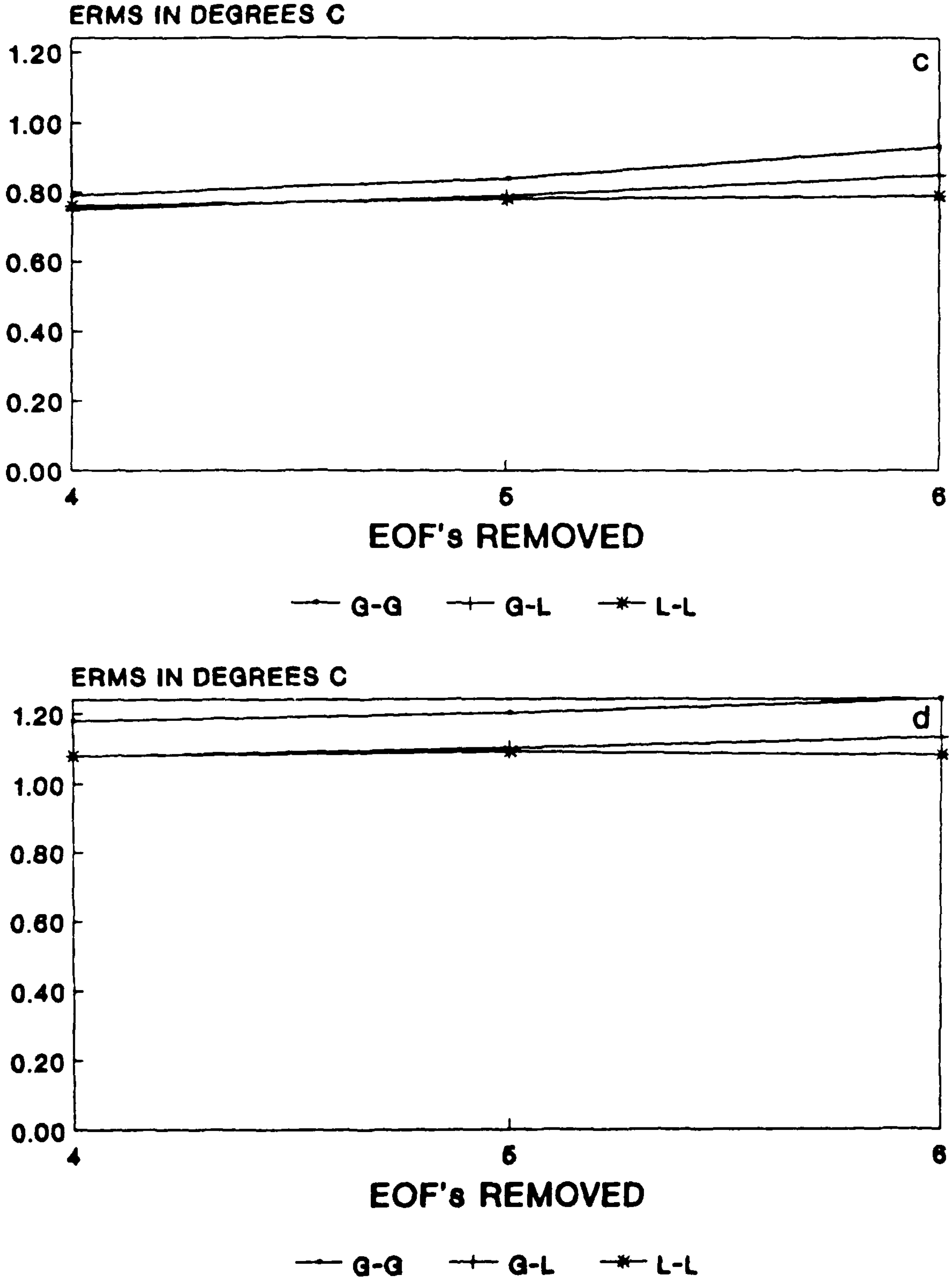

Fig. 2, continued.

RMS error is less when computed only over the ocean domain, and less still when the EOF's are defined in terms of data in that domain. The relative improvement in both instances becomes greater with increasing N (i.e., as fewer EOF's are retained), although the errors themselves of course become greater also. More improvement occurs in this case as a result of redefining the EOF's, as opposed to merely evaluating the errors over a different area.Figures 2b-d each show the same error curves as above for maximum retained bit values of (respectively) 11, 10, and 9. Comparing all four panels, one feature which is immediately apparent is that the increase in error with decreasing EOF retention is reduced as bit reduction becomes more severe, which was demonstrated previously for the case of globally-defined errors. Here we see that this result is true also for limited-area domains and different EOF definitions. Also evident is that the advantage of using the limited-area-defined EOF's diminishes for increasing bit truncation, although simply evaluating the errors over a more limited domain is still a source of improvement. Furthermore, error reduction in both instances (i.e., for both EOF definitions) is larger for greater N, and this difference becomes smaller as more bit reduction is employed. The magnitude of the improvement in RMS errors due to use of the limited domain is relatively small, although still significant, with maximum differences between adjacent curves of about 0.1 degrees C. The maximum total improvement (over the global error/global EOF case) is approximately twice this value.

Thus far we have concentrated on RMS errors in evaluating the horizontal domain influence, primarily because they give an integrated measure of error and are commonly examined in numerical weather prediction. However, the average absolute error and maximum absolute error were examined for these cases also, although we do not present them in a figure. As was discovered for globally defined errors, the behavior of average absolute error is very similar to that of RMS error, with smaller numerical values. Also as found previously, the maximum absolute error behaves somewhat differently from the integrated errors; in particular, as bit truncation is increased the results for locally defined EOF's tend to be worse than the corresponding results for global EOF's. There is still considerable improvement with respect to globally-evaluated errors, however.

A consideration of great importance is just what truncation provides the maximum acceptable error. For RMS error the largest acceptable value would probably be about 1 degree C, and preferably closer to 0.5 degree. In the N=4,5,6 cases examined here this condition is fulfilled clearly for the two instances of least bit truncation (i.e., the nonscaled and 11-bit maximum case), and somewhat less unambiguously for the case of 10 retained bits. Only the 9-bit case possesses RMS values greater than 1 degree. It is somewhat more difficult to specify a maximum acceptable

absolute error, but presumably one would not want one much larger than 5 degrees C. This requirement is met in all of our (N<7) cases except for the globally-evaluated error with 9 retained bits. Thus, the present results suggest that truncation to as few as 10 bits, keeping only 5 EOF's (out of 11), will yield errors which are acceptable, although this represents a maximum truncation (RMS error equal to 0.93 degrees C in the global case) and in a given instance greater accuracy may be desired.

Summarizing the results of this subsection, evaluating errors over a limited (rather than global) domain and defining vertical EOF's from data in that domain have both been shown to result in smaller domain-integrated errors. Maximum reductions in RMS error are about 0.1 degrees C for each of the two methods (if one characterizes regional evaluation of errors as a method), resulting in a total reduction over the global errors of 0.2 degrees C. The main source of error reduction is EOF redefinition in the case of mild bit truncation, and regional evaluation of errors in the case of severe bit truncation. Also, both sources yield larger error reductions as N, the number of EOF's removed, is increased. These results are consistent with the previously discovered effects of bit truncation, in that for a given bit scaling, EOF removal does not become significant until a certain threshold value is reached, this value increasing with increasing bit reduction. The increase in maximum absolute error with EOF redefinition (i.e., from global to local) in certain instances is more difficult to explain, and is likely related to the previously-mentioned fact that point measures of error will possess greater variability between cases. Overall, our findings suggest that a maximum truncation to 10 bits and 5 EOF's will give errors which are acceptable, although perhaps somewhat larger than would be desired in practice.

4.3 Spatial Characteristics of Errors

Because our previous analysis gives only information concerning domain-integrated (or extreme) errors, it is also of interest to examine the errors' three-dimensional spatial structure. Here we concentrate on the case with 5 retained EOF's and 10 retained bits. Examination of the global errors for globally-defined EOF's indicates a generally small-scale structure, consistent with the previously discussed tendency of our compaction method to eliminate small-scale information preferentially. Errors are concentrated mainly in the Southern Hemisphere, with the relative difference (i.e., compared to the Northern Hemisphere) being greatest at lower levels. The reason for this north-south asymmetry is unclear, but it may merely be an artifact of our particular case. However, comparative experiments (not discussed) demonstrate that, in this instance at least, the asymmetry is due almost totally to bit reduction rather than EOF truncation, thus implying a systematic reason for the bias. Studies with different cases will obviously be required to determine if this result is at all general. Other error characteristics are less easily summarized; that is, except for the above-

mentioned tendency of errors to predominate in the Southern Hemisphere, and to a lesser extent the occurrence of low-level error maxima over steep terrain, the error field appears basically random. (Interestingly, however, for EOF truncation in the absence of bit reduction, the errors are somewhat less randomly distributed, tending to concentrate more in middle and higher latitudes.) It should be noted that use of the locally-defined EOF's does not substantially change any of the preceding conclusions regarding the error field, although quantitative differences naturally exist.

Figure 3a and Fig. 3b each compare the total 850 mb temperature field of the control case with, respectively, the total 850 mb fields corresponding to the above truncation for globally-defined and limited area-defined EOF's. In both instances the errors, given by the difference between the solid (control case) and dashed (compacted case) lines, tend to be fairly small, with, as expected from previous results, greatest magnitude over the Southern Hemisphere. This can be seen more clearly in the error field (corresponding to Fig. 3a) itself, which we present as Fig. 4. The error field which is the counterpart to Fig. 3b appears very similar to Fig. 4 and is therefore not shown.

Concentrating on the North Pacific and North Atlantic regions, since these are the areas over which the limited-area EOF's are computed, it is seen from Fig. 3 that closer agreement with the control does in fact occur in the locally-defined EOF case. This is particularly evident in the North Pacific near 35N and the dateline, and in the North Atlantic over the region approximately bounded by 30W, 50W, 30N, 50N. Note that even in this instance, chosen because differences between the two schemes are most visible, the effect of varying the EOF definition is, over most areas, still rather subtle. Such a finding is consistent with previous results for RMS and other integrated errors in Section 4.2. Thus, computing vertical EOF's from data over a limited horizontal region, rather than over the whole globe, yields a slight positive impact on data compaction errors in the limited region, although this impact is generally most apparent for integrated measures and is less easily seen locally.

5. CONCLUSIONS

This study has demonstrated the feasibility of using both bit scaling and vertical EOF truncation to compress meteorological data. Based upon our results, a compaction factor of approximately two seems attainable from EOF truncation alone, with additional compression resulting from bit reduction to as few as 10 maximum bits. The amount of compaction in the latter instance is more difficult to express in terms of a factor, since it is precision-dependent. (For our choice of half precision, 32-bit words, the factor is approximately three.) Naturally, the above findings are crucially dependent on the accuracy which is desired. In our study, it is assumed that an RMS temperature error of no more than 1 degree C, and a maximum absolute error of no

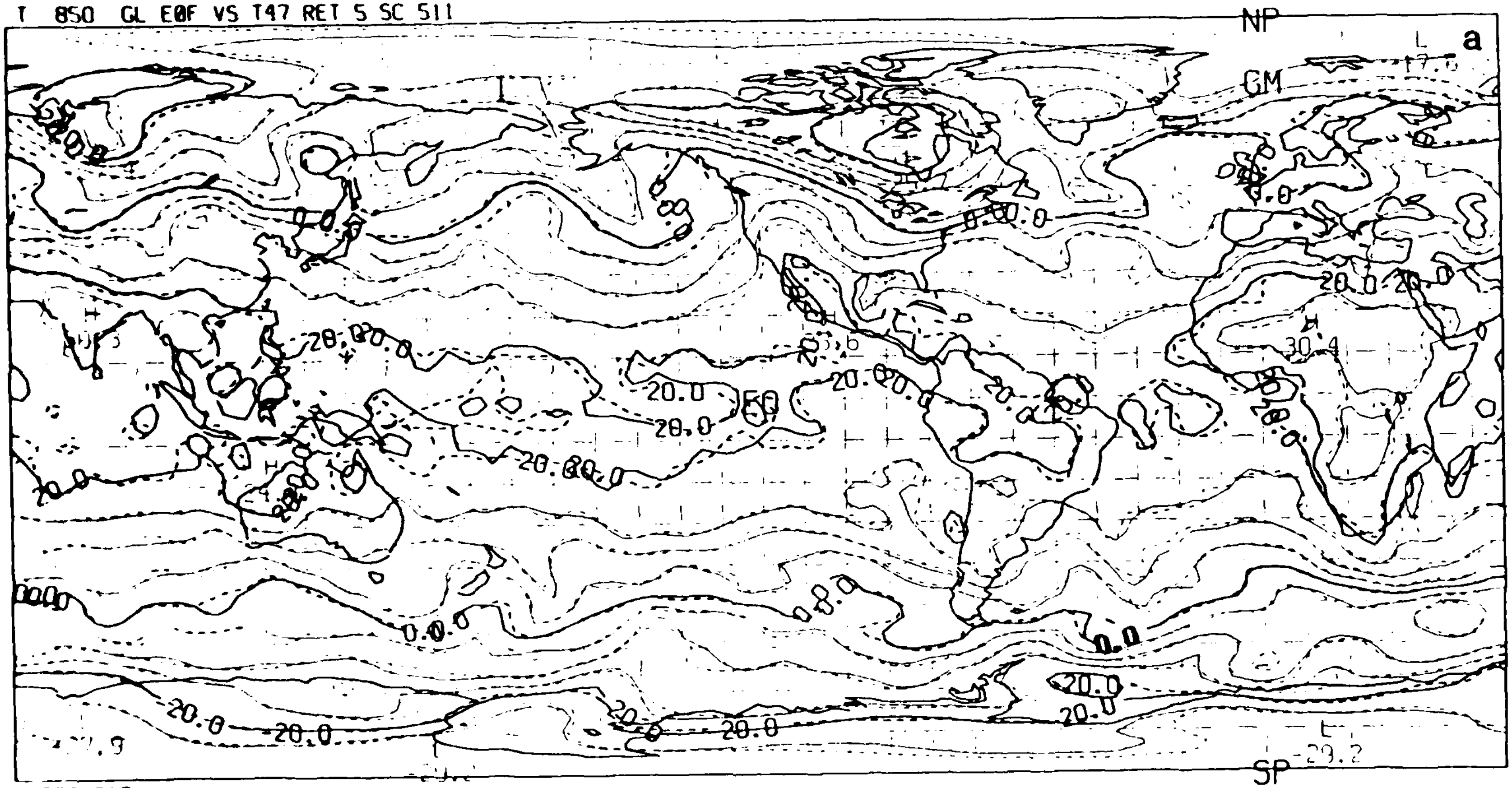

Fig 3. (a-b) Control temperature field at 850 millibars (solid), and (a) compacted 850 millibar temperature field using globally-defined EOF's (dashed), (b) compacted 850 millibar temperature field using EOF's defined from northern ocean data (dashed). Five EOF's and ten maximum bits are retained in computing the compressed field for both cases. Units: degrees Centigrade. Contour interval of 5 degrees C.

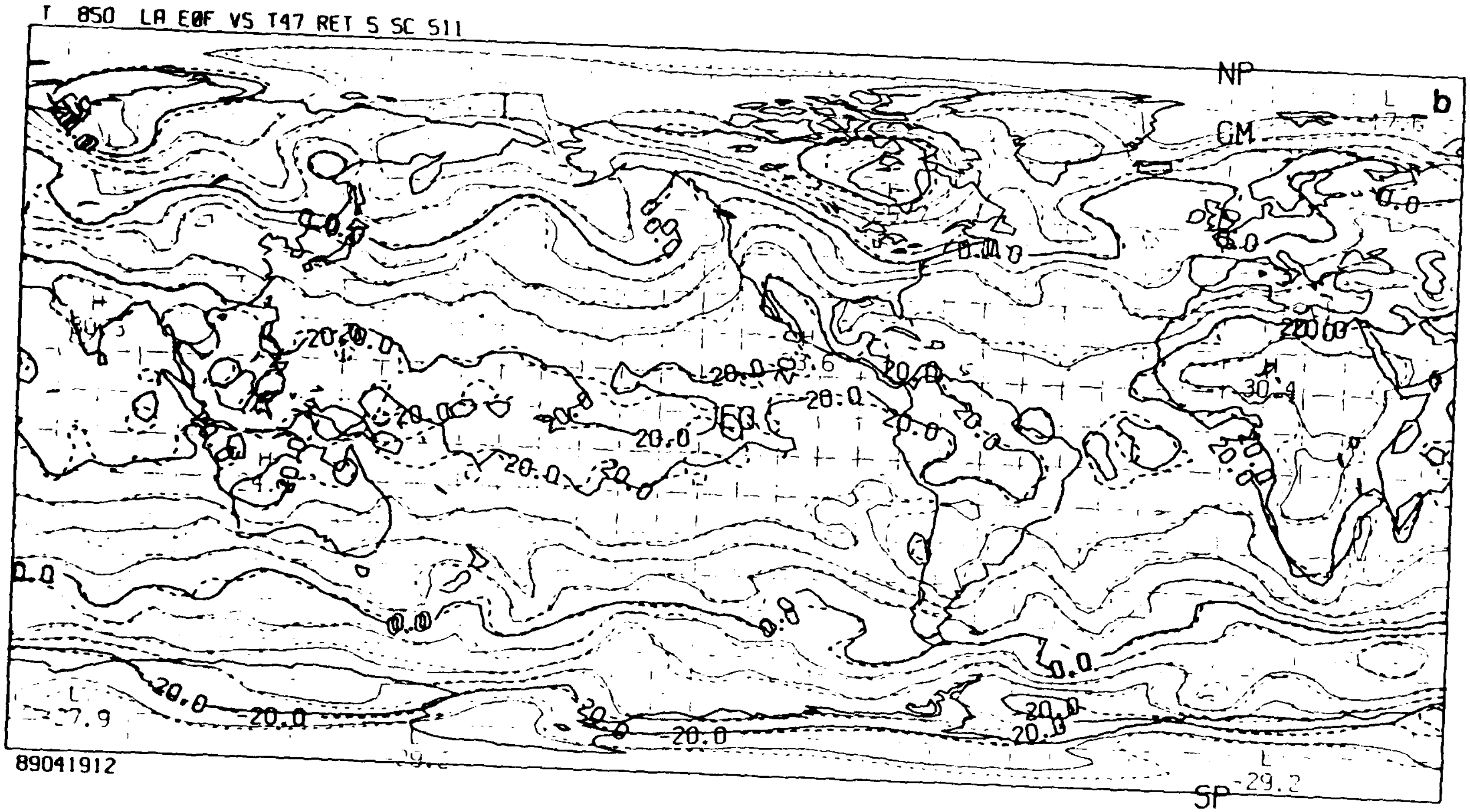

Fig. 3, continued.

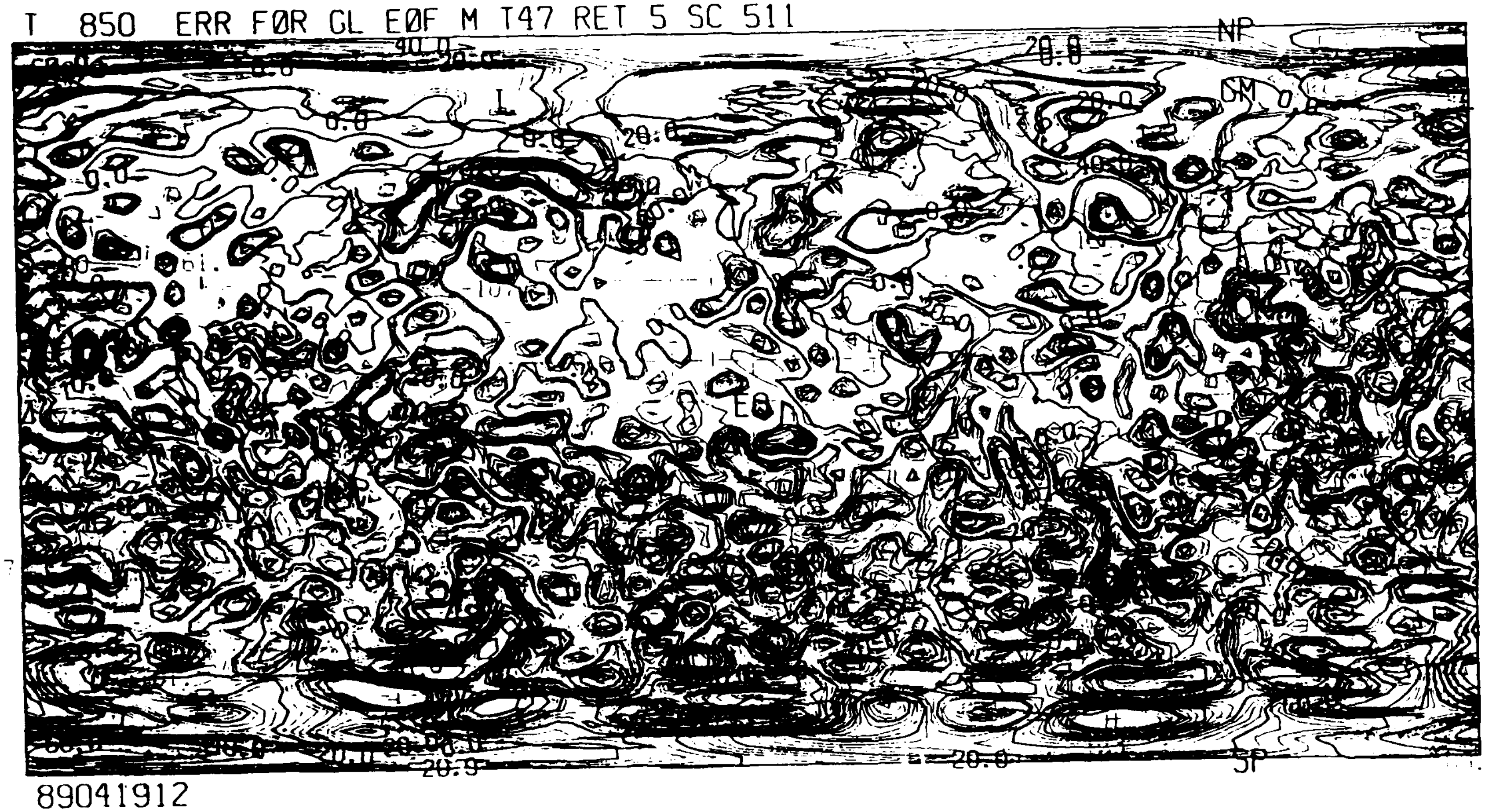

Fig. 4. Temperature error field, computed as the difference between the compressed field presented in Fig. 3a and the control field. Units: degrees Centigrade (plotted values have been multiplied by a factor of 50). Contour interval of 0.1 degrees C.

more than 5 degrees C, is permitted. For other particular applications this criterion may be either strengthened or relaxed, and the compression obtainable will vary accordingly. Additionally, there is the question of appropriate error tolerances for variables other than temperature.

For the global case, both domain-integrated and extreme errors increase (with a few minor exceptions in the latter instance) as fewer EOF's and fewer bits are retained; this obviously is as expected. A less obvious result is that, as bit truncation is increased, the <u>increase</u> in error due to EOF removal diminishes (to virtually zero at lower EOF truncations); the increase is present only once a threshold EOF truncation is reached. This threshold truncation becomes more severe with increasing bit reduction. The explanation for the above behavior is that, as previously discussed, one of the main effects of bit truncation is to set a number of the smaller spectral coefficients to zero, and, since the EOF's are ordered by increasing variance, there will exist certain EOF's for which <u>none</u> (or almost none) of the coefficients are nonzero. Truncating these coefficients obviously produces no effect. On the other hand, for the most severe EOF truncations (more severe than would ever be used operationally), the error is virtually independent of the number of maximum retained bits. One may therefore summarize by saying that bit reduction is the main cause of error for weak EOF truncation (i.e. most EOF's retained), and EOF removal is the main error source for strong EOF truncation. The contrasting behavior of errors due to the two compaction methods will be discussed in greater detail shortly. We note that for the amount of compaction likely to be employed in practice, <u>both</u> sources of error, bit reduction and EOF truncation, are important.

Evaluation of errors over the combined North Atlantic and North Pacific ocean regions, rather than over the entire globe, results in slightly better error statistics. This implies that more error in our compaction methods occurs over land (e.g., due to topography) than over the oceans, so that globally-averaged statistics may not be representative of the error over the regions of interest (which are, presumably, the oceans in most cases). The magnitude of the error difference in our study, however (approximately 0.1 degrees Centigrade RMS maximum) is rather small. Further improvement in error statistics occurs when the vertical EOF's are defined based upon data in only the limited horizontal region (i.e. the northern oceans). Here also, the improvement (maximum of 0.1 degrees RMS again) is modest. This improvement tends to decrease with increasing bit reduction and <u>decreasing</u> EOF truncation. Other characteristics of the limited area-evaluated errors (e.g., the functional dependence of the individual error curves upon bit scaling and EOF truncation) are very similar to characteristics of the global errors. It should be recalled that differences between cases with the two EOF definitions are most apparent from integrated statistics, such as RMS errors, and are not readily seen in contour plots except in a few instances. Nevertheless, the improvement, as measured by domain-averaged quantities, is definitely systematic.

The spatial structure of the compaction errors tends to be fairly random in most instances, except for the very noticeable tendency of errors to be concentrated in the Southern Hemisphere when bit truncation is present. In seeking the explanation for this phenomenon, one fact which should be considered is the small horizontal scale (in both directions) of the Southern Hemisphere errors; refer back to Fig. 4. Indeed, it may be shown that our bit reduction method, more so than EOF truncation, will tend to create small-scale errors. To understand why this is the case, it is necessary to recall on which combinations of horizontal and vertical structures the two compaction methods operate. The EOF truncation technique involves setting to zero all coefficients corresponding to a selected number of vertical EOF's; for a given EOF, however, the coefficients will tend to have a "red-noise" distribution in horizontal scale, so the error field, which of course possesses the same spectral characteristics as the neglected coefficients, will be essentially a sum of "red-noise" spectra. Therefore, assuming only limited cancellation of terms, the spectrum of the error field in this case should be weighted towards larger horizontal scales.

To determine the spectrum of the errors resulting from bit reduction, it is useful to think in terms of a "red-noise" spectrum in the vertical as well, since the EOF's may be ordered by decreasing variance; obviously, the analogy is inexact since the EOF's do not correspond to distinct vertical scales as such. If it is assumed that the effect of bit truncation is to zero out all coefficients with magnitude smaller than some value, then the eliminated coefficients will include those with large horizontal scale and small vertical variance, those with small horizontal scale and large vertical variance, together with a number which are intermediate between these two extremes. (Small-scale, small-variance coefficients will of course be excluded also.) Note that an "EOF cutoff" will exist, such that all EOF's with variance below a certain threshold have identically zero coefficients at all horizontal scales (this has been discussed previously); in addition, there will also be a "short-wave cutoff", that is, a horizontal wavelength such that all horizontal scales shorter than this wavelength have identically zero coefficients for all EOF's. If we then consider the total contribution of all eliminated coefficients for each horizontal scale, it is evident that the smallest scales will contribute most strongly, since they represent a sum over all EOF's, whereas the large-scale coefficients will be summed over only a few EOF's. (Note that we are here neglecting the effect of loss of accuracy for the coefficients which remain nonzero.) Having determined why bit reduction should preferentially favor small-scale errors, however, the question then becomes why the smallest scale errors are found predominantly in the Southern Hemisphere. Returning to an earlier speculation, the cause may be related to nothing more than the particular temperature structure of this one case. An additional possibility is that the Andes Mountains, which are of small zonal scale and thus not well represented in the NOGAPS spectral model, generate considerable small-scale noise; such a mechanism is consistent with the occurrence of the greatest north-south asymmetry in the error field at lower levels. This would be unlikely to impact the entire

hemisphere, however. As stated previously, experiments with other cases (and other meteorological variables) will be necessary to determine how general the above finding is, and to attempt to evaluate its cause.

6. RECOMMENDATIONS

Because this is a very preliminary study, involving only one case and one three-dimensional field, many of the recommendations necessarily consist of pointing out areas in which further investigation is required. One of the most obvious of these areas concerns extending the methods of this report to other cases, consisting of a variety of synoptic situations and seasons, and to meteorological variables other than temperature. This would help establish the generality of many of our conclusions, such as the degree of EOF truncation and bit reduction which yields acceptable errors. Also, our error tolerances were selected somewhat arbitrarily; future studies could attempt to define appropriate tolerances more rigorously (e.g., by considering the specific application of the compacted data), and for a more general set of meteorological variables.

In this study, defining the vertical EOF's based upon data only over the region of interest was demonstrated to give modest, although systematic, improvement in domain-integrated error statistics. Because this result is possibly dependent on the location and size of the domain, future experiments could be designed which would systematically examine the above dependence for a number of different regional domains. The influence of horizontal resolution on all of our results should also be considered, now that the T79 version of NOGAPS is operational.

One major improvement which could be made in our compaction technique involves the method of bit reduction, which as currently formulated results in the complete loss of much small-scale information. Two specific changes would help to alleviate this problem. The first change consists of expressing the complex spectral coefficients in terms of amplitude and phase, rather than in their original form, <u>then</u> scaling them. This is the method most frequently used in other compaction studies, and has as an advantage (among others) the absence of negative coefficient values. The other improvement possible would be to express the coefficients in terms of <u>deviations</u> from some background spectrum, as discussed in Section 2.2, then scale only the deviations. The background spectrum would be defined by only a few parameters, and so would not contribute significantly to the information needed to reconstruct the fields. This latter improvement should be particularly effective in helping to retain smaller-scale information, since the altered coefficients would all be approximately the same order of magnitude. (On the other hand, both the bit scaling and EOF truncation methods are designed to take advantage of the fact that the coefficients are <u>not</u> all of the same order of magnitude. Some compromise would therefore be required between this consideration and the desire to retain as much small-scale information as possible.)

An improvement which could be made regarding the definition of the EOF's would be to use a latitude-dependent weighting factor when computing them, in order to take account of the distortion arising from the latitude-longitude projection. Currently, when calculating EOF's from data over the whole globe, the North Pole and the South Pole each contribute 144 points, the same number as contributed by the entire equator and in fact by every other circle of latitude. The EOF's are therefore biased towards the atmospheric structure at high latitudes. Regionally computed EOF's of course possess this problem to a much lesser degree, although the effect still may not be negligible.

A question of great interest which should be addressed is the origin of the small-scale noise in the Southern, as opposed to Northern, Hemisphere. Experiments with different cases and variables would be useful in determining how general the occurrence of this phenomenon is: in particular whether, as previously suggested, it is merely a result of the particular date and meteorological field chosen. Also, it would be worthwhile to determine what impact different bit truncation schemes have on the appearance (or existence) of this noise, since in the present study it is the bit reduction, rather than EOF truncation, which is responsible for the small-scale error. If this error pattern is found to be general, and not just an artifact of a particular case or compaction technique, then diagnostic studies should be performed to determine its cause, since it might be a reflection of a deficiency in either the prediction model or the analysis scheme.

Most of the above recommendations consist of suggestions for future research, due to the nature of this study. The main recommendation with regard to the present time is that, since both of the compaction methods discussed here give significant amounts of data compression with reasonably small errors, they should be implemented, using regionally defined EOF's and (at least initially) truncation to 5 EOF's and 10 maximum bits, when an operational-quality code is available. It is thought such a code may be obtainable from the present software (i.e., that used for this report) without major modifications.

DISTRIBUTION LIST

COMMANDER
SPAWARSYSCOM
ATTN: CAPT HOFFMAN
WASHINGTON, DC 20361-5100

COMMANDER
SPAWARSYSCOM
ATTN: LCDR W. COOK
WASHINGTON, DC 20361-5100

COMMANDER
SPAWARSYSCOM
ATTN: CDR TRUMBOWER
WASHINGTON, DC 20361-5100

COMMANDER
SPAWARSYSCOM
ATTN: MR. D. MINTON
WASHINGTON, DC 20361-5100

COMNAVOCEANCOM
ATTN: CDR W. SHUTT
JCSSC, MS 39529-5000

COMNAVOCEANCOM
ATTN: R. BETCHER
JCSSC, MS 39529-5000

COMMANDING OFFICER
NAVOCEANO
ATTN: L. BERNARD
JCSSC, MS 39522-5001

NAVAL OCEANO. & ATMOS.
RESEARCH LABORATORY
ATTN: CODE 100
JCSSC, MS 39529-5004

NAVAL OCEANOGRAPHIC & ATMOS.
RESEARCH LABORATORY
ATTN: CODE 125L (10)
JCSSC, MS 39529-5004

COMMANDING OFFICER
FLENUMOCEANCOMCEN
ATTN: CDR ALDINGER
MONTEREY, CA 93943-5005

COMMANDING OFFICER
FLENUMOCEANCOMCEN
ATTN: M. PETERSON, 60A
MONTEREY, CA 93943-5005

CHIEF OF NAVAL OPERATIONS
NAVY DEPT. (OP-092X)
ATTN: MR. R. FEDEN
WASHINGTON, DC 20350-2000

OFFICE OF NAVAL RESEARCH
ATTN: HEAD, OCEAN SCI. DIV.
ARLINGTON, VA 22217-5000

OFFICE OF NAVAL TECHNOLOGY
ATTN: CODE 22
800 N. QUINCY ST.
ARLINGTON, VA 22217-5000

NAVWESTOCEANCEN
ATTN: OPERATIONS OFFICER
BOX 113
PEARL HARBOR, HI 96860

NAVEASTOCEANCEN
ATTN: OPERATIONS OFFICER
MCADIE BLDG. (U-117), NAS
NORFOLK, VA 23511-5399

NAVPOLAROCEANCEN
ATTN: OPERATIONS OFFICER
4301 SUITLAND RD.
WASHINGTON, DC 20395-5100

U.S. NAVOCEANCOMCEN
ATTN: OPERATIONS OFFICER
BOX 12, COMNAVMARIANAS
FPO SAN FRANCISCO 96630-2926

NAVAL OCEANOGRAPHIC & ATMOS.
RESEARCH LABORATORY
ATTN: CODE 300
JCSSC, MS 39529-5004

NAVPGSCOL
ATTN: OCEANOGRAPHY DEPT.
MONTEREY, CA 93943

NAVPGSCOL
ATTN: ACQUISITIONS
LIBRARIAN
MONTEREY, CA 93943

NAVAIRSYSCOM
ATTN: AIR-723D
WASHINGTON, DC 20361-0001

SPAWARSYSCOM
ATTN: CODE 312
NAT. CTR. #1
WASHINGTON, DC 20363-5100

SPAWARSYSCOM
ATTN: LCDR S. GRIGSBY
CODE PMW 145, NAVY DEPT.
WASHINGTON, DC 20363-5100

SPAWARSYSCOM
ATTN: CODE PMW 141-B1
WASHINGTON, DC 20363-5100

NAVOCEANSYSCEN
ATTN: J. RICHTER, CODE 54
SAN DIEGO, CA 92152-5000

D.W. TAYLOR RSCH. CENTER
ATTN: CODE 5220
BETHESDA, MD 20084-5000

NAVAL OCEANOGRAPHIC & ATMOS.
RESEARCH LABORATORY
ATTN: CODE 200
JCSSC, MS 39529-5004

NAVPGSCOL
ATTN: METEOROLOGY DEPT.
MONTEREY, CA 93943

NATIONAL METEOROLOGICAL CENTER
ATTN: NUMERICAL MODELING DIV.
WWB W32, RM 204
WASHINGTON, DC 20233

NCAR
ATTN: LIBRARY ACQUISITIONS
P.O. BOX 3000
BOULDER, CO 80302

U.S. NAVOCEANCOMCEN
ATTN: OPERATIONS OFFICER
BOX 31
FPO NEW YORK 09540-3200

U.S. NAVAL ACADEMY
ATTN: LIBRARY REPORTS
ANNAPOLIS, MD 21402

NATIONAL CLIMATIC CENTER
ATTN: L. PRESTON D542X2
FEDERAL BLDG. - LIBRARY
ASHEVILLE, NC 28801

COLORADO STATE UNIVERSITY
ATTN: ATMOS. SCI. DEPT.
FT. COLLINS, CO 80523

AFGWC/DAPL
ATTN: TECH. LIBRARY
OFFUTT AFB, NE 68113

AFGL/LY
ATTN: CODE LY
HANSCOM AFB, MA 01731

NATIONAL OCEANO. DATA CENTER
E/OC23, NOAA
ATTN: NUMERICAL MODELING DIV
WASHINGTON, DC 20235

WORLD METEOROLOGICAL ORG.
ATTN: GLOBAL MODELING UNIT
CASE POSTALE #5, CH-1211
GENEVA, SWITZERLAND

REPORT DOCUMENTATION PAGE		Form Approved OMB No. 0704-0188

1. Agency Use Only *(Leave blank).*

2. Report Date. May 1990

3. Report Type and Dates Covered. Final

4. Title and Subtitle.
Compaction of Global Data Fields

5. Funding Numbers.
Program Element No. 62435N
Project No. RM35G81
Task No. 1
Accession No. DN656756

6. Author(s).
A. H. Van Tuyl *

7. Performing Organization Name(s) and Address(es).
Naval Oceanographic and Atmospheric Research Laboratory
Atmospheric Directorate
Monterey, CA 93943-5006

8. Performing Organization Report Number.
NOARL Technical Note 27

9. Sponsoring/Monitoring Agency Name(s) and Address(es).
Office of Naval Technology
Arlington, VA 22217

10. Sponsoring/Monitoring Agency Report Number.
NOARL Technical Note 27

11. Supplementary Notes. Interagency agreement no. N6685687WR87041.
*The author is a visiting scientist from the National Center for Atmospheric Research (NCAR), Boulder, CO 80307-3000.

12a. Distribution/Availability Statement.
Approved for public release; distribution is unlimited.

12b. Distribution Code.

13. Abstract *(Maximum 200 words).*

Two methods of compacting global data fields are studied, both individually and in combination, and errors associated with the methods are systematically examined. The first scheme consists of expanding the data into an empirical orthogonal function (EOF) series in the vertical, then truncating the series at a selected number of terms. Because the EOFs are ordered by decreasing variance explained, this reduces the number of degrees of freedom while retaining most of the important vertical structure information. The second technique used is bit reduction, in which appropriately scaled data (here, spectral coefficients) are converted to integer form, with the scaling factor chosen so that the maximum data value is the largest integer expressible by some desired number of bits. Examination of compaction errors for various EOF truncations and bit scalings indicates that one important result of bit reduction is to set to zero all coefficients with magnitude below a certain threshold, causing EOF truncation up to a given point to have no impact on errors. Based on a somewhat arbitrarily selected maximum allowable RMS

((continued on reverse))

14. Subject Terms.
Vertical empirical orthogonal functions
Spherical harmonics
Data compaction
Bit reduction
NOGAPS

15. Number of Pages. 28

16. Price Code.

17. Security Classification of Report.	18. Security Classification of This Page.	19. Security Classification of Abstract.	20. Limitation of Abstract.
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	Same as report

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

Block 13, Abstract, continued.

temperature error of $1^{\circ}C$, a compaction factor of approximately two is obtainable from EOF truncation alone, and an additional factor of three from bit reduction (32 bits to 10), assuming half precision words. Future work is needed, however, to determine the generality of these results.